Published in final edited form as:
Comput Stat Data Anal. 2013 February 1; 58: 283–291. doi:10.1016/j.csda.2012.08.013.

Sample Size Calculation for Comparing Time-Averaged Responses in *K*-Group Repeated-Measurement Studies

Song Zhang and
Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX

Chul Ahn
Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX

Abstract

Many clinical trials compare the efficacy of K (≥3) treatments in repeated measurement studies. However, the design of such trials have received relatively less attention from researchers. Zhang & Ahn (2012) derived a closed-form sample size formula for two-sample comparisons of time-averaged responses using the generalized estimating equation (GEE) approach, which takes into account different correlation structures and missing data patterns. In this paper, we extend the sample size formula to scenarios where K (≥3) treatments are compared simultaneously to detect time-averaged differences in treatment effect. A closed-form sample size formula based on the noncentral χ^2 test statistic is derived. We conduct simulation studies to assess the performance of the proposed sample size formula under various correlation structures from a damped exponential family, random and monotone missing patterns, and different observation probabilities. Simulation studies show that empirical powers and type I errors are close to their nominal levels. The proposed sample size formula is illustrated using a real clinical trial example.

1 Introduction

Diggle et al. (2002) provided closed-form sample size formulas for clinical trials with repeated measurements, which compare the time-averaged responses and the rates of change based on a continuous outcome between two groups, assuming no missing data, the compound symmetry (CS) correlation among observations, and a balanced design. Jung & Ahn (2004) proposed a sample size formula to compare K-sample slopes in repeated-measurement studies using the generalized estimating equation (GEE) approach (Liang & Zeger 1986), which has been widely used to analyze repeated-measurement data due to its ability to accommodate missing values and robustness against misspecification of the true correlation structure. Time-averaged difference analysis is frequently used when the outcome varies with time (Zhang & Ahn 2011). For example, if mean blood pressure levels are compared between treatment groups by taking only one measurement from each subject, the experiment may have a poor performance due to substantial within-subject variation in blood pressure levels. Liu & Wu (2005) provided a sample size formula to test the time-averaged differences for unbalanced clinical trials between two treatment groups. Following

Correspondence should be sent to: Chul Ahn, Ph.D., Department of Clinical Sciences, UT Southwestern Medical Center, 5323 Harry Hines Blvd, E5.506, Dallas, TX 75390, Chul.Ahn@UTSouthwestern.edu.

the GEE approach, Zhang & Ahn (2012) extended the sample size calculation for time-averaged difference between two groups to allow for missing data, general correlation structures, and unbalanced randomization.

In this paper we further extend the approach of Zhang & Ahn (2012) to calculate sample sizes for repeated-measurement clinical trials where K (≥ 3) treatment groups are compared based on time-averaged differences. The derived sample size formula has a closed form, and is flexible enough to accommodate balanced or unbalanced experimental designs, arbitrary missing data patterns, and various correlation structures.

Suppose in a clinical trial enrolled subjects are randomly assigned to one of K treatment arms, and scheduled to be evaluated J times during the study period (at time $t_1, \cdots, t_J$). Here t_J indicates the end of study. Without loss of generality, we set $t_J = J$. We use y_{kij} to denote the continuous outcome variable obtained from the ith subject of the kth treatment arm at time t_j. The number of subjects within each treatment arm is denoted by n_k, and $n = \sum_{k=1}^{K} n_k$ is the total number of subjects. We define $r_k = n_k/n$ to be the probability of a patient being assigned to the kth treatment. To detect time-averaged differences among the K treatment arms, we consider the following model,

$$y_{kij} = b_k + \varepsilon_{kij}, \quad (1)$$

where b_k indicates a group-specific treatment effect and ε_{kij} is a zero-mean error term with variance σ^2. We assume ε_{kij} to be correlated within subjects, $\mathrm{Corr}(\varepsilon_{kij}, \varepsilon_{kij'}) = \rho_{jj'}$ with $\rho_{jj} = 1$, and independent across subjects. The null hypothesis of interest is $H_0 : b_1 = \cdots = b_K$.

First we briefly review the testing procedure without missing data. Based on an independent working correlation structure, the GEE estimator of b_k is

$$\widehat{b_k} = \frac{\sum_{i=1}^{n_k}\sum_{j=1}^{J} y_{kij}}{n_k J} = \frac{\sum_{i=1}^{n_k} \mathbf{1}'\boldsymbol{y}_{ki}}{\sum_{i=1}^{n_k} \mathbf{1}'\mathbf{1}}. \quad (2)$$

To facilitate a later extension that accommodates missing data, we present the estimator in a matrix form. Here $\boldsymbol{y}_{ki} = (y_{ki1}, \cdots, y_{kiJ})'$ is the vector of repeated measurements from the same patient and 1 is a vector with all J elements being 1.

Under the null hypothesis, we use b to denote the common value of treatment effects. i.e., $b_1 = b_2 = \cdots = b_K = b$. The GEE estimator of b is obtained by pooling observations from all K groups,

$$\widehat{b} = \frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l}\sum_{j=1}^{J} y_{lij}}{nJ} = \frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l} \mathbf{1}'\boldsymbol{y}_{li}}{\sum_{l=1}^{K}\sum_{i=1}^{n_l} \mathbf{1}'\mathbf{1}}.$$

We define a vector $\boldsymbol{B} = \sqrt{n}(\widehat{b_1} - \widehat{b}, \cdots, \widehat{b_{K-1}} - \widehat{b})'$. Plugging (1), we have

$$\widehat{b_k} - \widehat{b} = b_k - b + \frac{\sum_{i=1}^{n_k} \mathbf{1}'\boldsymbol{\varepsilon}_{ki}}{\sum_{i=1}^{n_k} \mathbf{1}'\mathbf{1}} - \frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l} \mathbf{1}'\boldsymbol{\varepsilon}_{li}}{\sum_{l=1}^{K}\sum_{i=1}^{n_l} \mathbf{1}'\mathbf{1}}.$$

Here $\boldsymbol{\varepsilon}_{ki} = (\varepsilon_{ki1}, \cdots, \varepsilon_{kiJ})'$. Under H_0, the central limit theorem suggests that as $n \to \infty$, vector $\boldsymbol{B}$ is approximately normal with a mean 0 and a $(K-1) \times (K-1)$ variance matrix $\boldsymbol{W} = (w_{kh})$. The value of w_{kh} is

$$w_{kh} = E\left[n \left(\frac{\sum_{i=1}^{n_k} \mathbf{1}'\boldsymbol{\varepsilon}_{ki}}{\sum_{i=1}^{n_k} \mathbf{1}'\mathbf{1}} - \frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l} \mathbf{1}'\boldsymbol{\varepsilon}_{li}}{\sum_{l=1}^{K}\sum_{i=1}^{n_l} \mathbf{1}'\mathbf{1}} \right) \cdot \left(\frac{\sum_{i=1}^{n_h} \mathbf{1}'\boldsymbol{\varepsilon}_{hi}}{\sum_{i=1}^{n_h} \mathbf{1}'\mathbf{1}} - \frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l} \mathbf{1}'\boldsymbol{\varepsilon}_{li}}{\sum_{l=1}^{K}\sum_{i=1}^{n_l} \mathbf{1}'\mathbf{1}} \right) \right]. \quad (3)$$

Note that we define $\boldsymbol{B} = \sqrt{n}(\widehat{b}_1 - \widehat{b}, \cdots, \widehat{b}_{K-1} - \widehat{b})'$ with length $(K-1)$ instead of $\sqrt{n}(\widehat{b}_1 - \widehat{b}, \cdots, \widehat{b}_{K-1} - \widehat{b}, \widehat{b}_K - \widehat{b})'$ with length K. The reason is that the K elements in the latter vector follow a linear constraint, $\sum_{l=1}^{K} r_l(\widehat{b}_l - \widehat{b}) = 0$, which in turn would lead to a singular variance matrix.

In practice $\boldsymbol{W}$ can be consistently estimated based on empirical error vectors $\hat{\boldsymbol{\varepsilon}}_{kij} = \boldsymbol{y}_{ki} - 1\hat{b}_k$. For hypothesis testing, we reject H_0 with type I error α if $\boldsymbol{B}'\hat{\boldsymbol{W}}^{-1}\boldsymbol{B} > \chi^2_{K-1,1-\alpha}$. Here $\chi^2_{K-1,1-\alpha}$ is the $100(1-\alpha)$th percentile of a χ^2 distribution with $(K-1)$ degrees of freedom.

2 Sample Size and Power Calculation

Suppose we would like to test the alternative hypothesis $H_a: b_1 = \theta_1, \cdots, b_K = \theta_K$, versus the null hypothesis $H_0: b_1 = \cdots = b_K$, based on statistic $\boldsymbol{B}'\hat{\boldsymbol{W}}^{-1}\boldsymbol{B}$. Under H_a, as $n \to \infty$, $\boldsymbol{B}'\hat{\boldsymbol{W}}^{-1}\boldsymbol{B}$ approximately has a noncentral χ^2 distribution with $K-1$ degrees of freedom and a noncentrality parameter $n\boldsymbol{\eta}'\boldsymbol{W}^{-1}\boldsymbol{\eta}$, where $\boldsymbol{W} = \lim_{n\to\infty} \hat{\boldsymbol{W}}$ and

$$\eta = (\theta_1 - \bar{\theta}, \cdots, \theta_{K-1} - \bar{\theta})'$$

with $\bar{\theta} = \sum_{k=1}^{K} r_k\theta_k$. Let $X^2_{K-1}(U)$ denote a noncentral χ^2 random variable with $K-1$ degrees of freedom and a noncentrality parameter U. Under type I error α and power $1-\beta$, the sample size is calculated by first solving for U from the following equation,

$$1 - \beta = P\left(X^2_{K-1}(U) > \chi^2_{K-1,1-\alpha}\right).$$

Denoting the solution by $U = U(K-1, \alpha, \beta)$, the required sample size is

$$n = \frac{U(K-1, \alpha, \beta)}{\eta'\boldsymbol{W}^{-1}\eta}. \quad (4)$$

The value of $U(K-1, \alpha, \beta)$ can be obtained through numerical search. The SAS function CNONCT also provides the solution. In the following we derive the expression of $\boldsymbol{W}$ or $\boldsymbol{W}^{-1}$ in the presence of missing data and arbitrary correlation structures, which eventually leads to a closed-form sample size formula. We assume that the outcomes are either measured at scheduled time $(t_1, \cdots, t_J)$ or missing, and the missing probabilities are only associated with time. Let δ_{kij} be an indicator which takes value 0/1 if a subject's outcome measurement at t_j is missing/observed. We define $p_j = E(\delta_{kij})$ to be the probability of a subject having an observation at time t_j, and $p_{jj'} = E(\delta_{kij}\delta_{kij'})$ to be the probability of a subject having observations at both t_j and $t_{j'}$. Note that $p_{jj} = p_j$.

Theorem 1. *As* $n \to \infty$, *the* $(K-1)\times(K-1)$ *variance matrix* $\hat{\boldsymbol{W}}$ *converges to*

$$\boldsymbol{W}=\frac{s}{\mu^2}\begin{pmatrix} (1/r_1-1) & -1 & -1 & \cdots & -1 \\ -1 & (1/r_2-1) & -1 & \cdots & -1 \\ -1 & -1 & (1/r_3-1) & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \cdots & (1/r_{K-1}-1) \end{pmatrix}.$$

Here $s=\sigma^2\sum_{j=1}^{J}\sum_{j'=1}^{J}p_{jj'}\rho_{jj'}$ *and* $\mu=\sum_{j=1}^{J}p_j$.

Proof. See Appendix A.1.

The special structure of $\boldsymbol{W}$ allows straightforward evaluation of its inverse, $\boldsymbol{W}^{-1}=\frac{\mu^2}{s}[\mathrm{diag}(\boldsymbol{r})+\mathrm{r}_{\mathrm{K}}^{-1}\boldsymbol{r}\boldsymbol{r}']$ with $\boldsymbol{r}=(r_1,\cdots,r_{K-1})'$. Then we have

$$\eta'\boldsymbol{W}^{-1}\eta=\frac{\mu^2}{s}\left[\sum_{k=1}^{K-1}r_k\eta_k^2+r_K^{-1}\left(\sum_{k=1}^{K-1}r_k\eta_k\right)^2\right]. \quad (5)$$

Plugging (5) into (4), we have the final sample size formula accounting for missing data in the test of time-averaged differences.

The required sample size is affected by missing data through observation probabilities (p_j and $p_{jj'}$) in s and μ. Different specifications of $p_{jj'}$ imply different missing patterns. For example, under random missing (RM), we have $p_{jj'}=p_jp_{j'}$. That is, having an observation at t_j is independent of having an observation at $t_{j'}$. On the other hand, under monotone missing (MM), a subject missing the measurement at t_j will miss all subsequent measurements. In this case we have $p_{jj'}=p_{j'}$ for $j'>j$. The sample size is also affected by within-subject correlation, which is characterized by $\rho_{jj'}$. Different correlation structures can be considered. The compound symmetry (CS) correlation structure assumes a constant correlation regardless of the temporal distance between two measurements, $\rho_{jj'}=\rho$ for $j\neq j'$. The autoregressive (AR) correlation structure, however, assumes $\rho_{jj'}=\rho^{|t_j-t_{j'}|}$. Thus measurements made close together will have a higher correlation than those made far apart. In this study we consider a flexible damped exponential family of correlations from Munoz et al. (1992). The within-subject correlation is parameterized by $\rho_{jj'}=\rho^{|t_j-t_{j'}|^{\phi}}$ with $0\le\phi\le 1$. Note that the CS and AR correlation structures are special cases of this family when ϕ takes value 0 and 1, respectively. With ϕ ranging between 0 and 1, we have a flexible model to describe the various correlation structures in real experiments.

Traditionally, researchers have taken a crude approach to adjusting sample size for missing data. Specifically, the observation probabilities usually decrease over time, $p_1\ge p_2\ge\cdots\ge p_J$, and we define $q=1-p_J$ to be the dropout rate at the end of study. To make adjustment for missing data, researchers first calculate the sample size under complete data ($p_1=p_2=\cdots=p_J=1$), denoted by n_0. Then the sample size with missing data is obtained by inflating n_0 by a factor of $1/(1-q)$. That is, $n^*=n_0/p_J$. This adjustment approach may be too conservative, resulting an unnecessarily large sample size and waste of resources. The reasons are: 1) Subject who dropped out at t_J might have partial observations (measurements at earlier times). These partial observations are involved in statistical inference but their contribution is ignored by the crude sample size adjustment. 2) The impact of missing patterns also needs

to be considered. Given the same dropout rate, observation probabilities that fall rapidly at the beginning of study and stabilize later will lead to more severe missingness than probabilities that drop steadily over time. Furthermore, given the same observation probabilities $\boldsymbol{P} = (p_1, \cdots, p_J)'$, the RM pattern suggests that missing values tend to be evenly distributed among all subjects while the MM pattern suggests that missing values tend to be concentrated among a subset of subjects. 3) The impact of missing data also depends on correlation. For example, if the repeated measurements are highly correlated, missing a few measurements may only lead to small information loss, and thus little impact on sample size.

The proposed sample size formula makes a better use of information. In $\mu = \sum_{j=1}^{J} p_j$, it considers all the observation probabilities over the study period instead of the dropout rate at the end of study only. In $s = \sigma^2 \sum_{j=1}^{J} \sum_{j'=1}^{J} p_{jj'} \rho_{jj'}$, the correlation ($\rho_{jj'}$) between pairs of measurements are taken into account as well as the probabilities that they are actually observed in the trial. The following theorem indicates that under realistic conditions, the proposed approach always leads to a saving in sample size compared with that based on the traditional adjustment for missing data.

Theorem 2. *In a clinical trial with repeated measurements, if a) the observation probabilities are non-increasing over time, $p_1 \geq p_2 \geq \cdots \geq p_J$; and b) the within-subject correlations are non-negative, $\rho_{jj'} \geq 0$; then we always have*

$$n \leq n_0 / p_J .$$

Furthermore, we have $n = n_0/p_J$ only under complete data, $p_1 = \cdots = P_J = 1$.

Proof. See Appendix A.2.

The two conditions are satisfied in most clinical trials. Given these two conditions, Theorem (2) is applicable regardless of missing pattern or correlation structure.

The sample size requirement also depends on the alternative hypothesis. One frequently assumed alternative hypothesis is that among the K groups, one receives a control treatment and the others receive different experimental treatments with similar efficacy. Without loss of generality, let θ_K denote the control treatment effect, and $\theta_1 = \cdots = \theta_{K-1} = \theta_K + \Delta$ denote the experimental treatment effects. This specification implies that $\bar{\theta} = \theta_K + (1 - r_K)\Delta$ and $\eta_k = r_K \Delta$ for $k = 1, \cdots, K - 1$. Then the sample size formula is

$$n = \frac{U(K-1, \alpha, \beta) s r_K}{\mu^2 \Delta^2 (1 - r_K) \sum_{k=1}^{K-1} r_k^2}.$$

Another widely used alternative hypothesis is that the experimental groups ($k = 1, \cdots, K - 1$) are ordered in treatment effect. For example, the dosage of an experimental drug increases over the groups, with a constant improvement in efficacy between consecutive dosages. In this case we have $\theta_k = \theta_K + k\Delta$ for $k = 1, \cdots, K - 1$. Here we only present the special case of a balanced design, $r_1 = \cdots = r_K = 1/K$. First we have $\bar{\theta} = \theta_K + \Delta(K - 1)/2$ and $\eta_k = \Delta[k - (K - 1)/2]$ for $k = 1, \cdots, (K - 1)$. The required sample size is expressed as

$$n = \frac{12 U(K-1, \alpha, \beta) s}{\mu^2 \Delta^2 (K^2 - 1)}.$$

3 Simulation

We conduct simulation studies to assess the performance of the proposed sample size. A clinical trial is conducted to compare the efficacy of $K = 4$ treatments based on time-averaged differences of $J = 6$ repeated measurements during the study period. The nominal power and type I error are set at $1 - \beta = 0.8$ and $\alpha = 0.05$, respectively. We explore the effects of a series of correlation structures from the damped exponential family with $\phi = 0$ (CS), 1/2, and 1 (AR(1)). Different values of parameter ρ are considered, ranging from 0.1, 0.25, to 0.5. We assess the impact of two missing patterns, RM and MM, with different trends in observation probabilities,

$\boldsymbol{P}_1 = (1, 0.82, 0.79, 0.76, 0.73, 0.7)$,

$\boldsymbol{P}_2 = (1, 0.94, 0.88, 0.82, 0.76, 0.7)$,

$\boldsymbol{P}_3 = (1, 1, 1, 0.9, 0.8, 0.7)$,

$\boldsymbol{P}_4 = (1, 1, 1, 1, 1, 1)$.

The first element of the four probability vectors are 1, indicating complete data at t_1. Vectors $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$ assume an equal dropout rate of 0.3 at the end of study, but via different paths. Specifically, $\boldsymbol{P}_1$ assumes a sharp drop at the early stage; $\boldsymbol{P}_2$ assumes a constant decrease throughout the study period; and $\boldsymbol{P}_3$ assumes missing data to only occur in the late stage. The scenario of complete data is described by $\boldsymbol{P}_4$. We consider two types of alternative hypotheses, both with $K - 1$ experimental treatments and a control treatment. The first one (denoted by $H_a^{(1)}$) assumes the $(K - 1)$ experimental treatments to have similar efficacy with $\Delta^{(1)} = 0.2$. The second one (denoted by $H_a^{(2)}$) assumes the $(K - 1)$ experimental treatments to be ordinal in efficacy with a constant improvement ($\Delta^{(2)} = 0.1$) upon one another. We specify variance parameter $\sigma^2 = 1$. For illustration purpose, we only present the results for balanced design $r_1 = \cdots = r_K = 1/K$. The sample size formula, however, is generally applicable to any randomization scheme.

A required sample size (n) is calculated for each combination of the aforementioned designing factors. Five thousand datasets, each containing n subjects, are then generated based on Model (1), with assumed missing patterns and correlation structures. We assess the empirical power and type I error in testing null hypothesis $H_0 : b_1 = \cdots = b_K$. The simulations are conducted using statistical software R 2.13.1 (R Foundation for Statistical Computing, Vienna, Austria). The R code is available upon request from the first author.

Tables 1 and 2 present estimated sample sizes, and empirical powers and type I errors, under alternative hypotheses $H_a^{(1)}$ and $H_a^{(2)}$, respectively. In both tables the empirical powers and type I errors are close to their nominal values, suggesting that the proposed sample size performs well. Furthermore, within each table, the sample sizes under $\boldsymbol{P}_4$ are exactly the same between the random and monotone missing patterns, because $\boldsymbol{P}_4$ represents complete data. We have a few observations. First, the within-subject correlation $\rho_{jj'} = \rho^{|t_j - t_{j'}|^\phi}$ is an increasing function of ρ and a decreasing function of ϕ. As the correlation ($\rho_{jj'}$) increases, the effective number of observations decreases, and the required sample size increases. Thus we observe in both tables that the sample sizes increase with ρ. Similarly, with ϕ increasing from 0 to 1, $\rho_{jj'}$ decreases as the correlation structure transforms from CS to AR(1), and the sample size requirement decreases. Under the monotone missing pattern, missing data tends to be concentrated within certain patients, which leads to a larger sample size requirement than that under the random missing pattern. The two tables also demonstrate the advantage of the proposed sample size in adjusting for missing data compared with the traditional approach. For example, under ($H_a^{(2)}$, $\rho = 0.1$, $\phi = 0$), the sample size under complete data is

219. To adjust for a dropout rate of 0.3, the crude approach would have required a sample size of $n^* = 219/0.7 = 313$, dramatically larger than those presented in Table 2: 255 for $\boldsymbol{P}_1$, 244 for $\boldsymbol{P}_2$, and 234 for $\boldsymbol{P}_3$ under RM; and 266 for $\boldsymbol{P}_1$, 250 for $\boldsymbol{P}_2$, and 236 for $\boldsymbol{P}_3$ under MM. The proposed sample size formula takes into account the strength of correlation (ρ), correlation structure (ϕ), trend in observation probability, and missing pattern, in the adjustment for missing data. This accurate adjustment enhances the efficiency in the use of medical resources in clinical trials.

Because in the damped exponential family, the correlation coefficients $\rho_{jj'}$ depend on both ϕ and ρ. To further investigate the association between correlation and sample size, we conduct an additional simulation study where we fix the value of $\bar{\rho}=(\sum_{k=2}^{K}\rho_{1k})/(K-1)$, the average of within-subject correlation coefficients. For ϕ=0, 1/2, and 1, we solve for the values of ρ such that $\bar{\rho}$ is fixed at particular levels. We explore three levels of $\bar{\rho}$: 0.1, 0.25, and 0.5. With (ϕ, ρ) obtained, we conduct simulations to assess the performance of the sample size formula. In Table 3 we present the simulation results for $H_a^{(1)}$. We can see that as ϕ increases, larger values of ρ are required to maintain the average level of correlation $\bar{\rho}$. The simulation results indicate that even for a given average level of correlation, different combinations of (ϕ, ρ) still lead to drastically different sample sizes. For example, under no missing data, if we fix $\bar{\rho}$ at 0.1, the required sample size is 364 under (ϕ = 0, ρ = 0.1) but 425 under (ϕ = 1, ρ = 0.33), a 17% difference. Thus the key factor affecting the sample size requirement is not the average correlation, but the combination of (ϕ, ρ). We reach similar conclusions from the simulation results for $H_a^{(2)}$ (the table omitted).

4 Real Data Example

For illustration, we apply the proposed sample size formula to a collaborative study of schizophrenia by the National Institute of Mental Health, which collected data on treatment-related changes in the overall severity (Hedeker & Gibbons 1997). A total of 437 patients were enrolled and randomly assigned to one of K = 4 medications arms: placebo, chlorpromazine, fluphenazine, and thioridazine. The latter three are anti-psychotic drugs. After initiating medication, patients were followed up weekly for 6 weeks. The outcome measurements followed a random missing pattern. The proportions of subjects with observations at weeks 1 to 6 are (0.98, 0.03, 0.86, 0.03, 0.02, 0.77). Thus only a small portion of subjects had measurements at weeks 2, 4, and 5. Because the three anti-psychotic drugs were considered to have similar effects, in a preliminary study, they were combined as one group and compared with the placebo. Model (1) with a CS correlation structure was fit to the data, and it was estimated that the time-averaged difference Δ = 0.99, variance σ^2 = 2.05, and correlation ρ = 0.45.

Suppose we would like to design a similar trial based on results from the aforementioned schizophrenia study. We set the levels of type I error at 0.05 and power at 0.90, and adopt a balanced design. Because most of the measurements were obtained at weeks 1, 3, and 6, we assume J = 3 and the vector of observation probabilities $\boldsymbol{P}$ = (0.98, 0.86, 0.77). If the three experimental treatments have a similar effect, it requires a total of n = 108 subjects to detect a time-averaged difference of Δ = 0.99. If there is no missing data, n = 101 subjects will be needed. By enrolling 437 subjects, the original study was powered to detect a time-averaged difference of Δ = 0.5. Finally, if we assume the effects of the three treatments to be ordinal at 0.79, 0.99, and 1.19 (the average remains to be 0.99), the estimated sample size will be n = 98.

5 Discussion

We have developed a closed-form sample size formula for a K-sample ($K \geq 3$) comparison of time-averaged responses that incorporates missing data, general correlation structures, and unbalanced randomizations. Our simulation results suggest that the proposed sample size formula performs well, with the empirical power and type I error close to the nominal levels under various correlation structures and missing data patterns. Our simulation results also demonstrate that correlation coefficients and correlation structures substantially influence the sample size requirement. The sample size under a CS structure is always larger than that under an AR(1) structure given the same ρ value. In the absence of information concerning the true correlation structure, a conservative approach would be to adopt the CS model.

The simulations in this paper show that with all other design parameters fixed, a larger correlation between repeated observations leads to a larger sample size. In Jung & Ahn (2004), which investigated the sample size calculation for comparing the rates of changes among K groups, a larger correlation between repeated measurements leads to a smaller sample size. Correlation affecting sample sizes differently in these two scenarios has been noted by Diggle et al. (2002) in two-sample comparisons.

We derived the sample size formula based on the assumption of missing completely at random. When the missing probability of a particular subjects depends on the subject's covariates, we can specify p_j as the average of individual probabilities over the distribution of the covariates in the population, and the results in this paper remain valid. If the missing probabilities depend on the subject's outcomes, an additional model for the missing mechanism is needed to achieve a valid statistical inference. This much more complicated model prevents the derivation of a closed-form sample size formula, and sample size requirement would have to be assessed through numerical studies. It is generally true, however, that a more complicated model would require a large sample size to maintain the level of power and type I error.

A heuristic approach for sample size calculation in K-sample trial has been to calculate the number of patients needed per treatment group using the sample size formula for 2-sample comparisons, and then multiply that number by K to obtain the total sample size needed for the study (Liu & Dahlberg 1995). Research is in progress to compare empirical powers between the heuristic and the proposed approaches under different scenarios such as (1) a control and $K - 1$ similar treatments, and (2) K ordered treatments.

Acknowledgments

This work was supported in part by NIH grants UL1TR0000451 and P30CA142543, and CPRIT grants RP110562-C1 and RP120670-C1.

Appendix A.1

Proof. With missing data, the expression of $\hat{b}_k$ is obtained from (2) by replacing **1** with $\boldsymbol{\delta}_{ki} = (\delta_{ki1}, \cdots, \delta_{kiJ})'$,

$$\widehat{b}_k = \frac{\sum_{i=1}^{n_k} \delta_{ki}' y_{ki}}{\sum_{i=1}^{n_k} \delta_{ki}' \delta_{ki}}.$$

The expression of $\hat{b}$, $\hat{b}_k - \hat{b}$, and w_{kh} in Section 1 can be modified similarly to accommodate missing data. To derive the expression of $\boldsymbol{W}$, first we have

$$w_{kh}=E\left[n\left(\frac{\sum_{i=1}^{n_k}\delta'_{ki}\boldsymbol{\varepsilon}_{ki}}{\sum_{i=1}^{n_k}\delta'_{ki}\delta_{ki}}-\frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\boldsymbol{\varepsilon}_{li}}{\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\delta_{li}}\right)\cdot\left(\frac{\sum_{i=1}^{n_h}\delta'_{hi}\boldsymbol{\varepsilon}_{hi}}{\sum_{i=1}^{n_k}\delta'_{hi}\delta_{hi}}-\frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\boldsymbol{\varepsilon}_{li}}{\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\delta_{li}}\right)\right].$$

We simplify the above expression using the fact that $\boldsymbol{\varepsilon}_{ki}$ are independent between subjects. When $k = h$,

$$w_{kk}=nE\left[\frac{\sum_{i=1}^{n_k}\delta'_{ki}\boldsymbol{\varepsilon}_{ki}\boldsymbol{\varepsilon}'_{ki}\delta_{ki}}{\left(\sum_{i=1}^{n_k}\delta'_{ki}\delta_{ki}\right)^2}+\frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\boldsymbol{\varepsilon}_{li}\boldsymbol{\varepsilon}'_{li}\delta_{li}}{\left(\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\delta_{li}\right)^2}-2\frac{\sum_{i=1}^{n_k}\delta'_{ki}\boldsymbol{\varepsilon}_{ki}\boldsymbol{\varepsilon}'_{ki}\delta_{ki}}{\left(\sum_{i=1}^{n_k}\delta'_{ki}\delta_{ki}\right)\left(\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\delta_{li}\right)}\right].$$

As $n\to\infty$, we have that $\sum_{i=1}^{n_k}\delta'_{ki}\boldsymbol{\varepsilon}_{ki}\boldsymbol{\varepsilon}'_{ki}\delta_{ki}$ converges to $nr_k s$, $\sum_{i=1}^{n_k}\delta'_{ki}\delta_{ki}$ converges to $nr_k\mu$, $\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\boldsymbol{\varepsilon}_{li}\boldsymbol{\varepsilon}'_{li}\delta_{li}$ converges to ns, and $\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\boldsymbol{\varepsilon}_{li}$ converges to $n\mu$. Thus

$$w_{kk}=ns\left(\frac{nr_k}{(nr_k\mu)^2}+\frac{n}{(n\mu)^2}-2\frac{nr_k}{(nr_k\mu)(n\mu)}\right)=\frac{(1-r_k)s}{r_k\mu^2}.$$

When $k\neq h$,

$$w_{kh}=nE\left[\frac{\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\boldsymbol{\varepsilon}_{li}\boldsymbol{\varepsilon}'_{li}\delta_{li}}{\left(\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\delta_{li}\right)^2}-\frac{\sum_{i=1}^{n_k}\delta'_{ki}\boldsymbol{\varepsilon}_{ki}\boldsymbol{\varepsilon}'_{ki}\delta_{ki}}{\left(\sum_{i=1}^{n_k}\delta'_{ki}\delta_{ki}\right)\left(\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\delta_{li}\right)}-\frac{\sum_{i=1}^{n_h}\delta'_{hi}\boldsymbol{\varepsilon}_{hi}\boldsymbol{\varepsilon}'_{hi}\delta_{hi}}{\left(\sum_{i=1}^{n_h}\delta'_{hi}\delta_{hi}\right)\left(\sum_{l=1}^{K}\sum_{i=1}^{n_l}\delta'_{li}\delta_{li}\right)}\right]. \quad (6)$$

Similarly it can be shown that w_{kh} converges to $-s/\mu^2$.

Thus we complete the proof of Theorem 1.

Appendix A.2

Proof. Under complete data ($p_1 = \cdots = p_J = 1$), we have $s_0=\sigma^2\sum_{j=1}^{J}\sum_{j'=1}^{J}=\rho_{jj'}$ and $\mu_0 = J$. Here we use subscript 0 to indicate their correspondence with n_0. Thus proving $n \le n_0/p_J$ is equivalent to proving that

$$\frac{n}{n_0/p_J}=\frac{\mu_0^2 p_J}{\mu^2}\cdot\frac{s}{s_0}=\frac{J^2 p_J}{\left(\sum_{j=1}^{J}p_j\right)^2}\cdot\frac{\sum_{j=1}^{J}\sum_{j'=1}^{J}p_{jj'}\rho_{jj'}}{\sum_{j=1}^{J}\sum_{j'=1}^{J}\rho_{jj'}}=\frac{p_J}{\overline{p}^2}\cdot\frac{J\overline{p}+2\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}p_{jj'}\rho_{jj'}}{J+2\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\rho_{jj'}}\le 1. \quad (7)$$

Here $\overline{p}=\sum_{j=1}^{J}p_j/J$.

Lemma 1. *Under Conditions a) and b),*

$$\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J} p_{jj'}\rho_{jj'} \le \overline{p}\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\rho_{jj'}.$$

Proof. Because $\rho_{jj'} \ge 0$ and, regardless of random or monotone missing pattern, $p_{jj'} \le p_{j'}$, we have

$$\frac{\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J} p_{jj'}\rho_{jj'}}{\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\rho_{jj'}} \le \frac{\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J} p_{j'}\rho_{jj'}}{\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\rho_{jj'}} = \frac{\sum_{j'=2}^{J}\left(p_{j'}\sum_{j=1}^{j'-1}\rho_{jj'}\right)}{\sum_{j'=2}^{J}\sum_{j=1}^{j'-1}\rho_{jj'}} = \sum_{j'=2}^{J} p_{j'}w_{j'}.$$

That is, the right hand side of inequality is a weighted average of $\{p_{j'}: j' = 2, \cdots, J\}$, with weights

$$w_{j'} = \frac{\sum_{j=1}^{j'-1}\rho_{jj'}}{\sum_{l=2}^{J}\sum_{j=1}^{l-1}\rho_{jl}}$$

With $\rho_{jj'} \ge 0$ from Condition b), we have $w_2 \le w_2 \le \cdots \le w_J$. Furthermore, Condition a) indicates that $p_1 \ge p_2 \ge \cdots \ge p_J$. In the weighted average, the weights decrease with the values of the elements. Thus

$$\sum_{j'=2}^{J} p_{j'}w_{j'} \le \overline{p}_{(-1)} \le \overline{p}.$$

Here $\overline{p}_{(-1)} = \sum_{j'=2}^{J} p_{j'}/(J-1)$ is the unweighted average of $\{p_{j'}: j' = 2, \cdots, J\}$. We have the last "≤" sign because $\bar{p}$ includes an additional element (p_1), which is no less than any of the elements in $\bar{p}_{(-1)}$. Thus we complete the proof of Lemma 1.

Using Lemma 1, the left hand side of inequality (7) is less than

$$\frac{p_J}{\overline{p}^2} \cdot \overline{p} = p_J/\overline{p} \le 1.$$

We have the last ≤ sign because p_J is the smallest element in $\boldsymbol{P} = (p_1, \cdots, p_J)$. Thus it is smaller than the average.

It is obvious from the above derivation that the equality sign only holds when $p_1 = \cdots = p_J = 1$.

References

Diggle, P.; Heagerty, P.; Liang, K.; Zeger, S. Analysis of longitudinal data. 2nd ed.. Oxford University Press; 2002.

Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychological methods. 1997; 2(1):64–78.

Jung S, Ahn C. K-sample test and sample size calculation for comparing slopes in data with repeated measurements. Biometrical Journal. 2004; 46(5):554–564.

Liang K, Zeger LL. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:45–51.

Liu H, Wu T. Sample size calculation and power analysis of time-averaged difference. Journal of Modern Applied Statistical Methods. 2005; 4(2):434–445.

Liu P, Dahlberg S. Design and analysis of multiarm clinical trials with survival endpoints. Controlled clinical trials. 1995; 16(2):119–130. [PubMed: 7789135]

Munoz A, Carey V, Schouten JP, Segal M, Rosner B. A parametric family of correlation structures for the analysis of longitudinal data. Biometrics. 1992; 48(3):733–742. [PubMed: 1420837]

Zhang S, Ahn C. How many measurements for time-averaged differences in repeated measurement studies? Contemporary Clinical Trials. 2011; 32(3):412–417. [PubMed: 21241827]

Zhang S, Ahn C. Sample size calculation for time-averaged differences in the presence of missing data. Contemporary Clinical Trials. 2012; 33(3):550–556. [PubMed: 22553832]

Table 1

Results of Simulation for $H_a^{(1)}$ with $\Delta^{(1)} = 0.2$

ρ	ϕ	P_1	P_2	P_3	P_4
(a) RM					
0.1	0	424(0.790, 0.064)	406(0.802, 0.058)	390(0.786, 0.058)	364(0.793, 0.052)
	$\frac{1}{2}$	362(0.812, 0.044)	345(0.788, 0.046)	331(0.804, 0.052)	302(0.797, 0.045)
	1	346(0.816, 0.056)	330(0.780, 0.046)	315(0.802, 0.068)	287(0.806, 0.050)
0.25	0	605(0.794, 0.058)	588(0.796, 0.046)	572(0.798, 0.060)	546(0.816, 0.039)
	$\frac{1}{2}$	483(0.766, 0.048)	468(0.820, 0.046)	455(0.808, 0.054)	425(0.816, 0.049)
	1	427(0.828, 0.050)	412(0.804, 0.052)	399(0.788, 0.058)	368(0.817, 0.045)
0.5	0	907(0.806, 0.054)	889(0.792, 0.052)	873(0.814, 0.028)	848(0.789, 0.046)
	$\frac{1}{2}$	753(0.794, 0.044)	739(0.808, 0.054)	727(0.796, 0.070)	696(0.792, 0.048)
	1	624(0.842, 0.052)	612(0.804, 0.060)	603(0.758, 0.064)	568(0.796, 0.068)
(b) MM					
0.1	0	443(0.814, 0.054)	416(0.812, 0.048)	393(0.814, 0.056)	364(0.793, 0.052)
	$\frac{1}{2}$	373(0.808, 0.034)	352(0.808, 0.056)	333(0.836, 0.056)	302(0.797, 0.045)
	1	355(0.782, 0.050)	335(0.798, 0.050)	317(0.764, 0.066)	287(0.806, 0.050)
0.25	0	654(0.774, 0.054)	612(0.804, 0.042)	579(0.796, 0.050)	546(0.816, 0.039)
	$\frac{1}{2}$	516(0.766, 0.062)	487(0.828, 0.050)	461(0.786, 0.050)	425(0.816, 0.049)
	1	452(0.776, 0.062)	427(0.782, 0.064)	405(0.816, 0.062)	368(0.817, 0.045)
0.5	0	1004(0.786, 0.070)	939(0.808, 0.064)	888(0.774, 0.046)	848(0.798, 0.046)
	$\frac{1}{2}$	831(0.770, 0.058)	781(0.824, 0.048)	740(0.779, 0.050)	696(0.792, 0.048)
	1	686(0.818, 0.078)	647(0.770, 0.052)	615(0.790, 0.050)	568(0.796, 0.068)

$H_a^{(1)}$ indicates the hypothesis that the $K - 1$ experimental treatments have similar efficacy with $\Delta^{(1)} = 0.2$ compared with the control treatment. ρ and ϕ are parameters in the damped exponential family. $\boldsymbol{P}_1 - \boldsymbol{P}_4$ represent different trends in observation probabilities. RM and MM indicates random and monotone missing patterns, respectively. The numbers in each cells are sample size (empirical power, empirical type I error).

Table 2

Results of Simulation for $H_a^{(2)}$ with $\Delta^{(2)} = 0.1$

ρ	ϕ	P_1	P_2	P_3	P_4
(a) RM					
0.1	0	255(0.808, 0.058)	244(0.814, 0.064)	234(0.830, 0.058)	219(0.812, 0.054)
	$\frac{1}{2}$	217(0.810, 0.072)	207(0.786, 0.050)	199(0.806, 0.080)	182(0.802, 0.056)
	1	208(0.800, 0.070)	198(0.816, 0.052)	189(0.808, 0.064)	172(0.799, 0.066)
0.25	0	363(0.796, 0.058)	353(0.814, 0.074)	343(0.800, 0.056)	328(0.792, 0.063)
	$\frac{1}{2}$	290(0.798, 0.050)	281(0.794, 0.054)	273(0.806, 0.056)	255(0.808, 0.075)
	1	256(0.810, 0.050)	247(0.820, 0.060)	240(0.806, 0.050)	221(0.785, 0.051)
0.5	0	545(0.814, 0.052)	534(0.802, 0.044)	524(0.796, 0.066)	509(0.793, 0.039)
	$\frac{1}{2}$	452(0.752, 0.036)	444(0.834, 0.060)	436(0.792, 0.058)	418(0.807, 0.060)
	1	375(0.818, 0.050)	368(0.798, 0.064)	362(0.808, 0.060)	341(0.807, 0.053)
(b) MM					
0.1	0	266(0.814, 0.056)	250(0.808, 0.050)	236(0.794, 0.052)	219(0.812, 0.054)
	$\frac{1}{2}$	224(0.802, 0.074)	211(0.806, 0.048)	200(0.818, 0.048)	182(0.802, 0.056)
	1	213(0.826, 0.072)	201(0.830, 0.070)	191(0.814, 0.058)	172(0.799, 0.066)
0.25	0	392(0.758, 0.044)	368(0.816, 0.050)	347(0.812, 0.048)	328(0.792, 0.063)
	$\frac{1}{2}$	310(0.834, 0.058)	292(0.842, 0.070)	277(0.826, 0.064)	255(0.808, 0.075)
	1	271(0.802, 0.064)	256(0.770, 0.048)	243(0.802, 0.070)	221(0.785, 0.051)
0.5	0	602(0.810, 0.056)	564(0.828, 0.056)	533(0.774, 0.050)	509(0.793, 0.039)
	$\frac{1}{2}$	499(0.830, 0.056)	469(0.798, 0.048)	444(0.802, 0.042)	418(0.807, 0.060)
	1	412(0.800, 0.048)	389(0.818, 0.052)	369(0.810, 0.050)	341(0.807, 0.053)

$H_a^{(2)}$ indicates the hypothesis that the $K - 1$ experimental treatments are ordinal in efficacy with a constant improvement ($\Delta^{(2)} = 0.1$) upon one another. ρ and ϕ are parameters in the damped exponential family. $\boldsymbol{P}_1 - \boldsymbol{P}_4$ represent different trends in observation probabilities. RM and MM indicates random and monotone missing patterns, respectively. The numbers in each cells are sample size (empirical power, empirical type I error).

Table 3

Results of Simulation for $H_a^{(1)}$ with $\Delta^{(1)} = 0.2$ given $\bar{\rho}$

$\bar{\rho}$	(ϕ,ρ)	P_1	P_2	P_3	P_4
(a) RM					
0.1	(0, 0.10)	424(0.790, 0.064)	406(0.802, 0.058)	390(0.786, 0.058)	364(0.793, 0.052)
	$(\frac{1}{2}, 0.22)$	458(0.801, 0.049)	443(0.810, 0.053)	429(0.797, 0.047)	400(0.821, 0.045)
	(1, 0.33)	483(0.798, 0.047)	469(0.812, 0.055)	457(0.808, 0.051)	425(0.811, 0.050)
0.25	(0, 0.25)	605(0.794, 0.058)	588(0.796, 0.046)	572(0.798, 0.060)	546(0.816, 0.039)
	$(\frac{1}{2}, 0.42)$	656(0.813, 0.058)	642(0.802, 0.044)	630(0.805, 0.049)	599(0.791, 0.049)
	(1, 0.57)	700(0.798, 0.046)	689(0.802, 0.047)	680(0.784, 0.043)	645(0.814, 0.051)
0.5	(0, 0.50)	907(0.806, 0.054)	889(0.792, 0.052)	873(0.814, 0.028)	848(0.789, 0.046)
	$(\frac{1}{2}, 0.65)$	958(0.806, 0.051)	944(0.806, 0.045)	931(0.812, 0.042)	901(0.783, 0.055)
	(1, 0.78)	1002(0.807, 0.063)	991(0.813, 0.065)	982(0.802, 0.042)	948(0.778, 0.046)
(b) MM					
0.1	(0, 0.10)	443(0.814, 0.054)	416(0.812, 0.048)	393(0.814, 0.056)	364(0.793, 0.052)
	$(\frac{1}{2}, 0.22)$	487(0.788, 0.043)	459(0.807, 0.051)	435(0.804, 0.058)	400(0.801, 0.055)
	(1, 0.33)	519(0.803, 0.055)	490(0.784, 0.049)	465(0.772, 0.064)	425(0.825, 0.063)
0.25	(0, 0.25)	654(0.774, 0.054)	612(0.804, 0.042)	579(0.796, 0.050)	546(0.816, 0.039)
	$(\frac{1}{2}, 0.42)$	719(0.804, 0.057)	676(0.801, 0.057)	641(0.794, 0.055)	599(0.809, 0.052)
	(1, 0.57)	775(0.812, 0.064)	731(0.828, 0.045)	695(0.811, 0.046)	645(0.788, 0.057)
0.5	(0, 0.50)	1004(0.786, 0.070)	939(0.808, 0.064)	888(0.774, 0.046)	848(0.798, 0.046)
	$(\frac{1}{2}, 0.65)$	1068(0.803, 0.054)	1002(0.799, 0.046)	949(0.787, 0.052)	901(0.821, 0.047)
	(1, 0.78)	1125(0.801, 0.054)	1058(0.817, 0.042)	1003(0.790, 0.068)	948(0.803, 0.042)

$H_a^{(1)}$ indicates the hypothesis that the $K-1$ experimental treatments have similar efficacy with $\Delta^{(1)} = 0.2$ compared with the control treatment. $\bar{\rho} = \sum_{k=2}^{K} \rho_{1k}/(K-1)$ is the average of within-subject correlation coefficients. (ρ,ϕ) are parameters in the damped exponential family that correspond to the particular levels of $\bar{\rho}$. $\boldsymbol{P}_1 - \boldsymbol{P}_4$ represent different trends in observation probabilities. RM and MM indicates random and monotone missing patterns, respectively. The numbers in each cells are sample size (empirical power, empirical type I error).